

## **EFFICIENT DOWNLOADING OF DOCUMENTS FROM THE INTERNET**

### **Field of the Invention**

The present invention relates to a system and method for the efficient downloading of information from the network, and in particular to a system and method for making use of the capacity of the network link which is left idle when a selected web page is being viewed.

### **Background of the Invention**

When one of today's network users uses a browse to show him an HTML document and, after a time, selects a link in the document which takes him to another HTML page, the browse does not begin to download the relevant data from the network until after the link has been selected. If the user has already looked at the page in question previously, and if as a result the page still happens to be in his computer's local cache, then it will be displayed more quickly. If however the page is not in a cache, the data contained in it will be downloaded from the server to the client via the network.

Over the period when the user is viewing a fully downloaded page, the limited transmitting capacity of the user's data link is not being used. Nevertheless, charges are incurred for the connection which has been established.

US patent 5896502 describes a method and system for the controlled transmission of a web page from a web server to a client system. The method is mainly directed to breaking off the

downloading of information from the network when the time taken by the transmission is of more than a defined length.

US 5931904 describes a method for the speedier display of information from the network by the installation of a local proxy.

US 5946697 describes a method for the compressed transmission of HTML pages.

WO 9908429A1 describes a system and method for the speedier downloading of the individual items of information making up an HTML document.

JP 10124413A describes a method for the prioritised downloading of components of the information making up an HTML document. The web author allots priorities reflecting the importance of the individual information objects forming an HTML document.

The object of the present invention is therefore to provide a system and method for the efficient downloading of information from the net where the downloading is adjusted to the user's actual behaviour.

### **Brief Description of the Invention**

In accordance with the present invention, the link information is expanded to include priority information, all the links featured on a web page are prioritised and, without being selected by the user, are automatically downloaded in the background in line with the prioritisation. This speeds up the downloading of a series of web pages to which there are connections by links. This means a considerable increase in the performance of the function for viewing linked web pages.

The automatic downloading takes place during the time when an already established connection exists. In this way any unused capacity of the connection is exploited to the full. This improves the economics or in other words allows fuller use to be made of the chargeable (telephone) connection which has been made to the network service provider.

The method according to the invention adjusts the automatic downloading to the user's habits: the downloading of subsequent pages can be influenced directly by setting certain configuring parameters and indirectly by analysing the user's behaviour during use (anticipatory downloading or preloading).

The method according to the invention makes it possible for web authors to predetermine the downloading of subsequent pages by assigning priorities. In addition, the user can himself make a selection determining the pages to be downloaded from configuration menus which may be part of a browse or of an add-on program for the relevant browsers.

#### **Brief Description of the Drawings**

The present invention will now be described by reference to a preferred embodiment and to figures, in which:

Fig.1 is a flow chart showing the method according to the invention,

Fig.2 shows the method according to the invention implemented in a client-proxy server architecture on the basis of a user configuration,

Fig.3 shows a further implementation of the method according to the invention in a client-proxy server architecture on the basis of a proxy server configuration;

Fig.4 shows a dialogue template for configuring the implementations shown in Figs. 2 and 3; and

Fig. 5 is a block diagram of a computer system and media that can be used with the present invention.

### Detailed Description of the Invention

To put the present invention into practice, the time during which the user is working his way through a page, or in other words the time during which the data link is not being used, is used to download those pages to which references are made by links. If one of the pages downloaded in anticipation is needed later on (namely when the user does in fact select the link concerned), it is already in the cache of the local computer and can be displayed at once.

The following mechanisms can be employed for the automatic anticipatory downloading:

1) Client-initiated automatic downloading:

- a) web-author-controlled downloading
- b) browser-controlled downloading
- c) user-controlled downloading

2) Server/gateway-initiated downloading:

- a) web-author-controlled downloading
- b) server-operator-controlled downloading

c) statistically controlled downloading.

Web-author-controlled downloading is achieved by establishing an interrelationship between web pages which belong together. The existing tags which define the links between web pages are expanded to include an additional parameter. In the tag used by HTML to make reference to another page, the author of the HTML page can state the priorities the various links are to be given, or in other words how important they are to be considered, and can say which of them are most likely to be pursued. The link with the lowest numbered priority is downloaded first during the "pause".

Example of an HTML page and its priority levels:

. . . . .

```
<a prio=5 href="/docs/gim/ocfgim.html"><b>General  
Information Web Document</b></a>
```

. . . . .

```
<a prio=6 href="/News/SystemTest/">OCF System Testing suite  
available</a>
```

. . . . .

```
<a prio=2 href='http://www.ibm.com/pvc=>IBM Pervasive  
Computing</a>
```

. . . . .

In the present example, the 'IBM Pervasive Computing' page would be downloaded first, followed by 'General Information Web' and then by 'OCF Testing suite available'. In this way, authors of web pages can greatly improve the overall impression the sites

make in terms of performance and can increase the acceptance of their pages.

In the case of browser-controlled downloading, the browse automatically creates user profiles which observe and analyse the behaviour of users. At the client end, the preferred solution includes in the browse a semantic network which is built up from information on the behaviour of the user by data mining and statistical methods. Then, the moment the user views a page, the browse prioritises the links included in the page on the basis of the information which has been assembled on the behaviour of the user and starts to download the most probable subsequent pages in background. Neuronal networks too may advantageously be used to detect behaviour. In this way the anticipatory downloading can be individually adjusted to the user's habits and can be optimised.

User-controlled downloading means that it is open to the user to determine the behaviour of the browse by using configuring menus and by setting options. The user can employ configuring parameters to specify whether it is complete pages, or alternatively only parts thereof, which are to be downloaded, in accordance with priorities or with probability. Lowest priorities/probabilities required for anticipatory downloading can be defined. Also, the user can define a series of pages which are to be downloaded automatically in the pauses which become available during use. A daily "internet round" can be defined in this way: the pages which have been defined, such as stock exchange bulletins, weather reports and newspaper headlines, will be downloaded continuously making full use of the network connection available and they can then be viewed in peace off-line without being connected to the network service provider.

Server-initiated downloading is preferably used in the area of ISDN or mobile telephony. In this case the gateway acts as an exchange between the terminals and the network. In the area of mobile telephony, the mobile phone communicates with the gateway server by WAP. When the user of the mobile phone wants information from the network, the gateway server makes the call to the desired web page on the network and downloads the page to its server and transmits the desired information from the web page selected to the mobile phone user. At the same time the gateway server performs the method according to the invention to identify and select links from the web page currently being processed and downloads the selected web pages to its cache in anticipation. This reduces the costly connection times between terminal users and the operator of the gateway which would be needed to call up information. The same method can also be employed in the ISDN area when the communication takes place via a gateway. It is also possible for the operator of the gateway server to employ a statistical process for determining the relevant web page by using a data mining program in order to select the web page which is to be downloaded in anticipation. Similarly simple operator configuration settings, e.g. sequential downloading, may also be considered.

Fig.1 is a flow chart showing the method according to the invention.

An network URL (universal resource locator) is entered to select a given web page on the network (step 101) and the site is downloaded to the client's volatile or non-volatile cache (step 102). There the web page is checked for all the links it includes (step 103). This check is made by means of an add-on program, e.g. a so-called plug-in, or a browse extension. The job of the add-on program or browse extension is to identify links by reference to predefined settings and download them

automatically. There are various methods which can be employed to identify and select the links. The links themselves may contain information, e.g. priority information. The add-on program or browse extension reads the items of priority information in the individual links and downloads the respective web pages from the server to the client in the sequence determined by the order of priorities while the user is still looking at the original web page.

However, for this to be possible it is essential for the web author to have prepared the links by providing them with priority information. Where this has not been done, other methods have to be employed to select the links. These methods may for example be:

Sequential downloading of the links - In this case the links are identified and automatically downloaded sequentially by the add-on program or browse extension.

Sequential downloading of the links in accordance with a user setting - The settings in this case may for example be these: "from centre", "top-down" or "bottom-up".

Determination of behaviour-specific parameters and allocation of links in the light of them. This method usually requires the use of a data mining program.

Downloading of the links in accordance with search terms which have been entered and which are freely definable by the user.

The web pages which can be addressed by the links are downloaded automatically to the client by the add-on program or browse extension. Depending on the browse setting the downloading may



be to the cache of the RAM (memory cache) or the cache of the hard disk (disk cache).

When a user selects a new link on the web page and enables it (step 104), the add-on program or browse extension checks to see whether the web page allocated to the link is already in store in the client's cache (step 105). If it is, it is downloaded from the cache (step 106) and displayed. The new web page too is checked for any links it may contain by the add-on program or browse extension. If it does contain links, the method according to the invention is started again.

If however the link which the user has selected and enabled is one whose web page is not in store in the client's cache, the web page concerned has to be downloaded from the network in a fresh operation (step 107). The method described above for identifying the links a page has then starts for the new web site.

Fig.2 shows the method according to the invention implemented in a client-proxy server architecture on the basis of a user configuration.

The client-proxy server architecture comprises a client having a browse and a cache, and a proxy server having a cache. The client communicates with the network via the proxy server. What is stored in the client's cache is data representing web pages/links which have been downloaded in the past 201.

Stored in the proxy server's cache are web pages which have been downloaded in anticipation 202. These web pages are selected by means of a data mining program 203 and a user-set configuration 204. The data mining program has access to the data in the client's cache in this case. On the basis of this data and the

user-set configuration, certain links are selected from those present on a web page which the client is currently dealing with and their associated web pages are downloaded to the proxy server's cache in accordance with the priorities assigned to them. If the user selects a link, the web page associated with the link is transmitted from the proxy server to the client. This causes the method to be reinitiated, i.e. the data mining program and the user-set configuration select certain links on the web page which is currently in use and automatically download the web pages assigned to these links to the proxy server's cache.

Fig.3 shows a modified version of the client-proxy server architecture shown in Fig.2. In this implementation, the web pages to be downloaded are selected by the operator of the proxy server 301. Where the proxy server is being operated by a company, it is the company that creates a proxy configuration 302 on the basis of its operating requirements. An additional data mining module 303 can change the proxy configuration. The data mining module is preferably installed on the proxy server. What the proxy server configuration lays down in this case is a definition of priority criteria. These priority criteria are accepted by a program (preload module) installed on the proxy server and are compared with the information which is supplied by the data mining program to the proxy server. Working from the priority criteria and the information provided by the data mining program, the preload module selects the appropriate web pages which are to be preloaded, i.e. downloaded in anticipation. This is done as shown in Fig.2.

Fig.4 shows an example of a dialogue template for configuring the implementations shown in Figs. 2 and 3.

The user can preferably fill in the dialog template in the preset sequence.

Where the user has selected sequential downloading, the downloading takes place without any selection of the content which is downloaded. However, the sequential downloading can be acted on by means of the parameters, "from the centre", "top-down" and "bottom-up".

Where the links to the web pages already contain priority information, the user can assign a lowest priority. In the example shown, this is a priority of 2. All links down to a priority of 2 are downloaded in anticipation.

Finally, the user can select behaviour-specific priorities. As part of this he can set a lowest probability. If the links identified do not meet this lowest probability requirement, they are ignored. The possibility also exists of selecting "changeover probabilities" and "page-content probabilities". As well as this, the user can select "standard priorities" by entering a code word.

The priorities selected can be re-arranged into a sequence relative to one another. On the right of the dialog template shown as an example, the selection options are prioritised as follows:

Priority 1 has the priority defined at the server end. On the web page being viewed, all those subsequent links will be downloaded in anticipation which have already been given an HTML tag of "Prio=1" at the server end. With this configuration, "Prio=2" would be ignored.

Priority 2 have all the subsequent links which have the code word "Smartcard".

Under priority 3, the data miner determines which subsequent links are most probable. No lowest probability has been selected.

Changeover probabilities are cases where, for example, when somebody is on a corporate web site, he will often want to change over to look at the share prices quoted for the company as well.

Page-content probabilities cause the data miner to take account of whether the description included in the link mentions current favourite subjects.

Priority 4 is like priority 1 except that links marked Prio=2 are also included.

As shown in Fig. 5, the software for performing the functions of the present invention can be provided, or the results received, from a computer system 502 and placed on a computer useable media 504, such as an optical or magnetic media, and can be displayed on a computer responsive display system 506.

It should be apparent that a number of changes, substitutions and alterations can be made to what has been described. Therefore, it should be understood that the present invention is not limited to what has been described but includes those embodiments within the scope and spirit of the appended claims.